VOLUME 32 · NUMBER 6 · FEBRUARY 20 2014

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Nomogram to Predict Cycle-One Serious Drug-Related Toxicity in Phase I Oncology Trials

David M. Hyman, Anne A. Eaton, Mrinal M. Gounder, Gary L. Smith, Erika G. Pamer, Martee L. Hensley, David R. Spriggs, Percy Ivy, and Alexia Iasonos

See accompanying editorial on page 489

David M. Hyman, Anne A. Eaton, Mrinal M. Gounder, Erika G. Pamer, Martee L. Hensley, David R. Spriggs, and Alexia Iasonos, Memorial Sloan-Kettering Cancer Center; David M. Hyman, Mrinal M. Gounder, Martee L. Hensley, David R. Spriggs, and Alexia Iasonos, Weill Cornell Medical College, New York, NY; and Gary L. Smith and Percy Ivy, National Cancer Institute, Bethesda, MD.

Published online ahead of print at www.jco.org on January 13, 2014.

D.M.H. and A.A.E. contributed equally to this work.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: David M. Hyman, MD, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, New York, NY 10065; e-mail: hymand@mskcc.org.

0732-183X/14/3206w-519w/$20.00

DOI: 10.1200/JCO.2013.49.8808

ABSTRACT

Purpose
All patients in phase I trials do not have equivalent susceptibility to serious drug-related toxicity (SDRT). Our goal was to develop a nomogram to predict the risk of cycle-one SDRT to better select appropriate patients for phase I trials.

Patients and Methods
The prospectively maintained database of patients with solid tumor enrolled onto Cancer Therapeutics Evaluation Program–sponsored phase I trials activated between 2000 and 2010 was used. SDRT was defined as a grade ≥ 4 hematologic or grade ≥ 3 nonhematologic toxicity attributed, at least possibly, to study drug(s). Logistic regression was used to test the association of candidate factors to cycle-one SDRT. A final model, or nomogram, was chosen based on both clinical and statistical significance and validated internally using a bootstrapping technique and externally in an independent data set.

Results
Data from 3,104 patients enrolled onto 127 trials were analyzed to build the nomogram. In a model with multiple covariates, Eastern Cooperative Oncology Group performance status, WBC count, creatinine clearance, albumin, AST, number of study drugs, biologic study drug (yes *v* no), and dose (relative to maximum administered) were significant predictors of cycle-one SDRT. All significant factors except dose were included in the final nomogram. The model was validated both internally (bootstrap-adjusted concordance index, 0.60) and externally (concordance index, 0.64).

Conclusion
This nomogram can be used to accurately predict a patient's risk for SDRT at the time of enrollment. Excluding patients at high risk for SDRT should improve the safety and efficiency of phase I trials.

J Clin Oncol 32:519-526. © 2014 by American Society of Clinical Oncology

INTRODUCTION

The primary objective of phase I studies is to define the maximum-tolerated doses (MTDs) of novel drugs or combinations in a specific patient population. Selecting the most appropriate patients to participate in these important studies represents a significant challenge. Strict eligibility criteria requiring good performance status and near-normal organ function (bone marrow, kidney, and liver) are used to maximize patient safety. Although these eligibility criteria prevent 30% of potential patients from participating in phase I trials,[1,2] many patients still experience progression or die early in the course of phase I trials.[3] Several prognostic models, including the Royal Marsden Hospital and European prognostic scores, can help identify patients with the poorest survival.[3-7] However, these existing prognostic models do not predict which patients are at increased risk for serious drug-related toxicity (SDRT) in phase I trials.[8,9]

Patients who are symptomatic from their cancer, prior treatments, and medical comorbidities are almost certainly at increased risk for toxicity, but data quantifying these relationships are sparse. Because SDRTs that occur during cycle one typically qualify as dose-limiting toxicities (DLTs), it is important we understand which patient characteristics influence SDRT risk. Eastern Cooperative Oncology Group performance status (ECOG PS) has consistently been shown to be a predictor of SDRT in phase I studies in models with multiple explanatory

variables.[8,9] Older age, liver metastasis, low WBC count, elevated platelet count, and elevated bilirubin may increase SDRT risk, but there is disagreement over their predictive power.[8,9] To date, there are no tools available to estimate the risk of SDRT in phase I studies.

Enrolling patients highly susceptible to SDRTs at low dose levels has detrimental effects on the safety, speed, and efficiency of phase I studies. SDRTs that occur in high-risk patients typically require early dose expansion to prove the dose level is safe before dose escalation can be continued. This increases the number of patients needed to determine the MTD and adds substantially to the cost of phase I studies.

We conducted a multi-institutional pooled analysis of National Cancer Institute Cancer Therapeutics Evaluation Program (CTEP) –sponsored phase I studies throughout North America, with the goal of developing a clinical prediction tool, or nomogram, that can estimate a patient's risk for developing an SDRT at the time of enrollment. This tool can aid in clinical decision making regarding patient enrollment onto phase I studies.

PATIENTS AND METHODS

Study Design and Patient Eligibility

A multicenter cohort of patients treated in CTEP-sponsored[10] phase I trials activated between 2000 and 2010 was used for model derivation. Data were provided from the Clinical Trials Monitoring System (CTMS) database, which is managed by Theradex Systems (Princeton, NJ). The CTMS database is prospectively maintained, with robust data management and auditing practices.[11]

Trials evaluating either cytotoxic agents or molecularly targeted agents (defined as drugs that target an extra- or intracellular mechanism different from those associated with conventional chemotherapy such as DNA, tubulin, or cell-division machinery) alone and in combination were included. Trials involving vaccines, radiation therapy, and locoregional therapies were excluded. Dedicated organ dysfunction trials were also excluded. Eligible patients were adults (age $\geq$ 18 years), had solid tumors, and had received at least one dose of study drug(s). Patients were also required to meet generally accepted phase I laboratory criteria as follows: absolute neutrophil count $\geq 1 \times 10^9$/L, hemoglobin $\geq$ 8 g/dL, platelet count $\geq 75 \times 10^9$/L, AST/ALT $\leq 5\times$ upper limit of normal, and total bilirubin $\leq 2\times$ upper limit of normal. Patients with leukemia and lymphoma and those with incomplete data for $\geq$ one of the covariates included in the final nomogram were excluded. All patients had regular follow-up visits as specified by the protocol onto which they were enrolled.

An independent cohort of patients consecutively enrolled onto phase I trials between 2009 and 2012 in the Developmental Therapeutics Clinic at Memorial Sloan-Kettering Cancer Center was used as a validation set. CTEP-sponsored studies included in the derivation set were excluded. Patients and trials in the validation set were required to meet the same eligibility criteria as the derivation set.

Outcome

The primary outcome was the presence or absence of any SDRT during cycle one for each patient. An SDRT was defined as a grade $\geq$ 3 nonhematologic or grade $\geq$ 4 hematologic toxicity attributed as possibly, probably, or definitely related to study treatment; this was done to: one, use a uniform outcome definition across all trials, and two, mirror the definition of a DLT used by most phase I trials. Grade 3/4 electrolyte disorders without associated symptoms (ie, clinically insignificant toxicities) were excluded from the SDRT definition. For each protocol, cycle one was defined from the date of first drug administration plus the protocol-defined cycle length. All toxicities were graded using Common Terminology Criteria for Adverse Events (version 4).[12] Toxicity data collected on trials using older Common Terminology Criteria for Adverse Events versions were mapped to version 4 using published National Cancer Institute conversion tables.[13]

Candidate Factors

The following categories of candidate factors were considered: one, commonly used phase I eligibility criteria; two, measures of prior treatment exposure; three, measures of disease burden; four, protocol factors known or suspected to influence toxicity; and five, other factors with a clinically plausible relationship to drug toxicity susceptibility.

Model Building and Validation

A detailed description of the methodology used to build and validate the model is provided online (Appendix, online only). Briefly, analyses were conducted using logistic regression.[14] The Breslow-Day test for homogeneity of the odds ratio confirmed that the association between candidate covariates and SDRT was similar across agent classes (Appendix Table A1, online only).[15] Therefore, one model for all agent types combined was pursued. The final regression model was chosen based on the clinical and statistical significance of the predictors, following previously published methodology.[16] For each patient, the predicted probability of an SDRT in cycle one was calculated using the final logistic regression model underlying the nomogram. The model was internally validated in the derivation set by using 500 bootstrap samples to estimate the bias-corrected concordance index (C-index), a measure of the predictive accuracy of the model. The C-index was also calculated in the independent (ie, external) validation set. The calibration of the nomogram, which measures how far predictions are from observed outcomes, was assessed via a calibration plot.[17]

RESULTS

Patient and Trial Characteristics of Derivation Set

Data on 3,104 patients treated in 127 phase I trials were analyzed. Baseline patient characteristics are listed in Table 1. A broad range of tumor types was represented. Median age was 58 years (range, 18 to 87 years). The median number of prior systemic therapies was five (range, zero to 33). The median number of metastatic sites was three (range, zero to 17).

Forty-three percent of patients were treated in trials of molecularly targeted agents, 19% in trials of cytotoxics, and 38% in trials combining cytotoxics and molecularly targeted agents. Thirty-seven percent of patients received one study drug, and 63% of patients received $\geq$ two study drugs. Twenty-one percent of patients were treated at the highest administered dose level for the particular study onto which they were enrolled. Dose level was expressed relative to the highest dose administered in each protocol using a four-tier categorical variable such that highest was the highest dose, highest -1 was the second-highest dose tested in each protocol, and so on.

SDRT Rate During Cycle One

A total of 1,040 SDRTs occurred in 728 unique patients during cycle one, yielding an overall serious toxicity rate of 23.5%. Thirteen patients (0.4%) died as a result of drug-related toxicity during cycle one. SDRT rates varied across agent classes, with rates of 20.4% for cytotoxics, 22.2% for molecularly targeted agents, and 26.4% for a combination of both agent types. Table 2 lists the SDRTs observed by drug class and toxicity category. The most common SDRTs were hematologic (observed in 7.5% of patients), GI (6.0%), constitutional (5.0%), and metabolic (4.8%).

Factors Associated With SDRT

Patients with multiple SDRTs during cycle one contributed only once in this analysis. The results of a univariable analysis of the

Table 1. Baseline Patient Demographic and Clinical Characteristics (N = 3,104)

Characteristic	No.	%
Primary tumor site		
GI	1,069	35
Genitourinary	373	12
Thoracic	370	12
Breast	350	11
Gynecologic	295	10
Sarcoma	242	8
Head and neck	197	6
Melanoma and skin	164	5
Brain and unknown	44	1
Sex		
Male	1,539	50
Female	1,565	50
Age, years		
Median	58	
Range	18-87	
ECOG performance status		
0	892	29
1	2,056	66
≥ 2	156	5
No. of prior systemic therapies		
0-2	860	28
3	358	12
≥ 4	1,881	60
Missing	5	0
Prior radiation therapy		
Yes	1,446	47
No	1,653	53
Missing	5	0
No. of metastatic sites		
0	147	5
1	530	17
2	592	19
3	581	19
≥ 4	1,020	32
Missing	234	8
Metastatic site*		
Lung	1,109	36
Liver	1,090	35
Lymph node	612	20
Bone	217	7
Brain	8	0
Missing	234	8
Sum of longest tumor dimensions, cm		
Median	8.1	
Range	0-49.5	
Laboratories		
WBC, × 10^9/L		
Median	6.7	
Range	2.1-38.2	
ANC, × 10^9/L		
Median	4.6	
Range	1.1-34.9	
ALC, × 10^9/L		
Median	1.9	
Range	0.2-21.6	
Hemoglobin, g/dL		
Median	12.2	
Range	8-17.3	

(continued in next column)

Table 1. Baseline Patient Demographic and Clinical Characteristics (N = 3,104) (continued)

Characteristic	No.	%
Platelets, × 10^9/L		
Median	255	
Range	78-1,114	
Albumin, g/dL		
Median	3.8	
Range	1.8-5	
AST, U/L		
Median	27	
Range	7-176	
ALT, U/L		
Median	23	
Range	2-170	
Total bilirubin, mg/dL		
Median	0.5	
Range	0.1-1.9	
Alkaline phosphatase, U/L		
Median	102	
Range	25-1,915	
LDH, U/L		
Median	231	
Range	29-10,405	
Creatinine clearance, mL/min†		
Median	93	
Range	13-125	
Baseline symptoms (grade ≥ 2)*		
Constitutional	546	18
Cardiovascular	521	17
GI	309	10
Neurologic	138	4
Hematologic	438	14
Study drug agent class		
Molecularly targeted drug	1,345	43
Cytotoxic	575	19
Cytotoxic and molecularly targeted drug	1,184	38
Biologic study drug		
Yes	172	6
No	2,932	94
Trial eligibility		
Broad	2,851	92
Disease specific	253	8
No. of study drugs		
1	1,137	37
≥ 2	1,967	63
Study drug dose level		
Highest	640	21
Highest −1	900	29
Highest −2	549	18
≤ Highest −3	1,015	32

Abbreviations: ALC, absolute lymphocyte count; ANC, absolute neutrophil count; ECOG, Eastern Cooperative Oncology Group; LDH, lactate dehydrogenase.
*Patients may fall into more than one category for these covariates.
†Estimated by Cockcroft-Gault equation, capped at 125 mL/minute.

association of patient and study factors to cycle one SDRTs are summarized in Table 3. Significant factors ($P \leq .10$) included ECOG PS (0 *v* 1 *v* ≥ 2), WBC count, creatinine clearance, albumin, AST, ALT, total bilirubin, alkaline phosphatase, number of study drugs (one *v* ≥ two), biologic study drug, study drug class, dose level, and baseline constitutional symptoms (grade 0 *v* 1 *v* ≥ 2).

Table 2. Serious Drug-Related Toxicities During Cycle One (N = 3,104)

Toxicity	Overall		Cytotoxic		Combination		Molecular	
	No. of Patients	Rate per 1,000 Patients (%)	No. of Patients	Rate per 1,000 Patients (%)	No. of Patients	Rate per 1,000 Patients (%)	No. of Patients	Rate per 1,000 Patients (%)
Total	728	234.5	117	203.5	313	264.4	298	221.6
Constitutional (total)	155	49.9	26	45.2	57	48.1	60	53.5
Fatigue	108	34.8	22	38.3	38	32.1	48	35.7
Other	53	17.1	5	8.7	21	17.7	27	20.1
Cardiovascular (total)	93	30.0	10	17.4	23	19.4	60	44.6
Hypertension	37	11.9	1	1.7	0	0.0	36	26.8
Other	57	18.4	9	15.7	23	19.4	25	18.6
GI (total)	187	60.2	19	33.0	91	76.9	77	57.2
Anorexia	26	8.4	2	3.5	8	6.8	16	11.9
Diarrhea	78	25.1	7	12.2	48	40.5	23	17.1
Nausea/vomiting	93	30.0	12	20.9	35	29.6	46	34.2
Other	32	10.3	2	3.5	14	11.8	16	11.9
Respiratory	20	6.4	3	5.2	7	5.9	10	7.4
Renal	10	3.2	0	0.0	1	0.8	9	6.7
Metabolic (total)	149	48.0	18	31.3	62	52.4	69	51.3
Liver abnormalities	51	16.4	5	8.7	17	14.4	29	21.6
Amylase/lipase elevation	9	2.9	0	0.0	0	0.0	9	6.7
Hyperglycemia	45	14.5	3	5.2	26	22.0	16	11.9
Other	51	16.4	10	17.4	23	19.4	18	13.4
Dermatologic	44	14.2	0	0.0	9	7.6	35	26.0
Neurologic	17	5.5	1	1.7	8	6.8	8	5.9
Hematologic (total)	233	75.1	60	104.3	127	107.3	46	34.2
Neutropenia	226	72.8	60	104.3	122	103.0	44	32.7
Anemia	2	0.6	0	0.0	1	0.8	1	0.7
Thrombocytopenia	8	2.6	2	3.5	4	3.4	2	1.5
Other (total)	66	21.3	15	26.1	27	22.8	24	17.8
Thrombosis/hemorrhage	26	8.4	3	5.2	13	11.0	10	7.4
Death	13	4.2	4	7.0	5	4.2	4	3.0
Infection	28	9.0	8	13.9	11	9.3	9	6.7
Miscellaneous	6	1.9	1	1.7	3	2.5	2	1.5

NOTE. Toxicities within each major category and subcategory and total toxicities are listed on per-patient basis. Because one patient may have experienced multiple toxicities within a category, and category totals only allow each patient to contribute one toxicity, subcategories may sum to number greater than category total.

Nomogram Development and Internal Validation

A nomogram was built accounting for both statistical and clinical significance. The multiple covariate model results for two candidate models (with and without dose level included) are summarized in Table 4. Because dose level was expressed relative to the highest dose administered in each trial, this covariate cannot be known at the time of patient enrollment. As a result, dose level was not included in the final nomogram, which is shown in Figure 1. The nomogram demonstrated good accuracy for predicting cycle-one SDRT in the derivation set, with an unadjusted C-index of 0.61 (95% CI, 0.59 to 0.63) and a bootstrap-corrected C-index of 0.60. Calibration curves for the nomogram in the derivation set are shown in Figure 2 and suggest excellent model calibration, with model estimates being close to observed rates. The distribution of model-estimated risk is available online (Appendix Fig A1, online only).

Independent Nomogram Validation

The final nomogram was also validated in an independent data set (n = 234; Appendix Table A2, online only). Patient factors were comparable between the validation and derivation cohorts. The validation cohort had a higher proportion of trials involving one study drug (68% *v* 37%) and biologic therapy (31% *v* 6%) compared with the derivation cohort, in addition to a lower SDRT rate (13.3% *v* 23.5%). When the nomogram was used to predict cycle-one SDRTs in the validation cohort, the C-index was 0.64 (95% CI, 0.53 to 0.75). The median model-predicted probability of SDRT was 17.2% (range, 7.3% to 34.5%).

DISCUSSION

In this report, we investigated the relationship of a number of baseline patient and study characteristics with the chance of developing early SDRTs. We present an internally and externally validated nomogram that uses information available before patient enrollment to estimate the risk of cycle-one SDRT, an outcome that mirrors the definition of DLT used by most phase I studies. This nomogram demonstrates that some patients in phase I trials are at high risk for developing early SDRTs, regardless of the dose of study drug they receive. In our derivation cohort, 15.5% of patients had an estimated risk for SDRT during cycle one of at least 30%, invariant to dose. Most physicians and regulatory agencies consider an SDRT rate of ≥ 30% to be unacceptably high and use this threshold to define the MTD in phase I trials.[19,20] This observation calls into question the safety of including

Table 3. Univariable Analysis of Predictors of Cycle-One Serious Drug Related Toxicities

Factor	OR	95% CI	*P*
Sex			.997
Male	1.00	0.85 to 1.18	
Female	Ref		
Age (10-year increase)	1.04	0.97 to 1.12	.218
ECOG performance status			.005
0	Ref		
1	1.15	0.95 to 1.40	
≥ 2	1.86	1.29 to 2.69	
No. of prior systemic therapies			.652
0-2	Ref		
3	1.14	0.86 to 1.52	
≥ 4	1.02	0.84 to 1.23	
Prior radiation therapy			.778
Yes	1.02	0.87 to 1.21	
No	Ref		
No. of metastatic sites			.306
0	Ref		
1	1.42	0.90 to 2.23	
2	1.08	0.69 to 1.69	
3	1.16	0.74 to 1.82	
≥ 4	1.18	0.76 to 1.81	
Sum of longest tumor dimensions, cm*	—		.556
Laboratories			
WBC, $\times 10^9$/L*	—		.073
ANC, $\times 10^9$/L*	—		.529
ALC, $\times 10^9$/L*	—		.382
Hemoglobin, g/dL*	—		.425
Platelets, $\times 10^9$/L*	—		.170
Albumin, g/dL*	—		.001
AST, 20-μg/mL increase	1.17	1.09 to 1.27	< .001
ALT, U/L*	—		.036
Total bilirubin, mg/dL*	—		.002
Alkaline phosphatase, U/L*	—		.089
LDH, U/L*	—		.497
Creatinine clearance, 10-mL/min increase†	0.96	0.93 to 0.99	.021
Hemoglobin, g/dL			.268
8-9	1.47	0.74 to 2.92	
9-10 g	1.19	0.56 to 2.55	
> 10	Ref		
Study drug agent class			.006
Molecularly targeted drug	Ref		
Cytotoxic	0.90	0.71 to 1.14	
Cytotoxic and molecularly targeted drug	1.26	1.05 to 1.52	
Hyperglycemia at baseline, CTCAE grade			.219
0	Ref		
1	1.14	0.51 to 2.57	
≥ 2	2.06	0.93 to 4.56	
Constitutional symptoms at baseline, CTCAE grade			.059
0	Ref		
1	1.15	0.95 to 1.39	
≥ 2	1.33	1.05 to 1.70	
GI symptoms at baseline, CTCAE grade			.805
0	Ref		
1	0.94	0.78 to 1.13	
≥ 2	1.01	0.76 to 1.34	
Pain symptoms at baseline			.802
No	Ref		
Yes	0.96	0.69 to 1.34	
Biologic study drug			.001
Yes	0.51	0.33 to 0.80	
No	Ref		
Trial eligibility			.249
All solid tumors	Ref		
Disease Specific	0.83	0.61 to 1.14	
No. of study drugs			< .001
1	Ref		
≥ 2	1.65	1.37 to 1.97	
Study drug dose level			< .001
Highest	2.29	1.82 to 2.90	
Highest −1	1.51	1.21 to 1.89	
Highest −2	1.62	1.26 to 2.08	
≤ Highest −3	Ref		

Abbreviations: ALC, absolute lymphocyte count; ANC, absolute neutrophil count; CTCAE, Common Terminology Criteria for Adverse Events; ECOG, Eastern Cooperative Oncology Group; LDH, lactate dehydrogenase; NA, not applicable; OR, odds ratio; Ref, referent.

*Restricted cubic splines; OR not applicable.

†Estimated by Cockcroft-Gault equation, capped at 125 mL/minute.

high-risk patients in phase I studies, as well as the reproducibility of the toxicity data these patients contribute to characterizing the safety profile of a new anticancer agent. These results have potentially important implications for the design and conduct of phase I studies.

This nomogram could be used to further inform decision making and allow both patients and physicians to carefully weigh the risks of participation against the potential benefits of the experimental drug(s). To illustrate how this nomogram might be used to aid decisions, we can consider two hypothetic patients with advanced solid tumors being considered for a study with two nonbiologic drugs; both patients meet all standard phase I eligibility criteria. The first patient has an ECOG PS of 1, WBC count of 3.5 $\times 10^9$/L, albumin of 3.0 g/dL, AST of 50 units/dL, and creatinine clearance of 65 mL/min. The second patient has an ECOG PS of 0, WBC count of 10×10^9/L, albumin of 3.8 g/dL, AST of 24 units/dL, and creatinine clearance of 85 mL/min. Using the proposed nomogram, the first patient has a predicted SDRT risk of 44%, and the second patient has one of 22%. The high predicted risk for the first patient may dissuade the physician from recommending, or the patient from participating in, this protocol. This example also illustrates the important role the number of study drugs and agent type and class can have on expected toxicity. For example, if these same two patients were treated in a protocol with a single biologic drug, the predicted a priori risk of an SDRT in the first and second patients would drop to 20% and 8%, respectively. These data suggest that the eligibility criteria for a phase I trial could be calibrated to the anticipated toxicity of the study being planned.

Our analysis provides further evidence that the factors that predict for patient survival and drug toxicity in phase I trials may only partially overlap.[21] Most prognostic models include either a direct or indirect measure of tumor burden, such as number of

Table 4. Multiple-Covariate Models of Predictors of Cycle-One Serious Drug-Related Toxicities (N = 3,104)

	Final Nomogram			Nomogram Plus Dose		
Significant Factor Included in Nomogram	OR	95% CI	*P*	OR	95% CI	*P*
ECOG performance status			.0237			.0175
0	Ref			Ref		
1	1.10	0.90 to 1.33		1.11	0.91 to 1.36	
2	1.72	1.17 to 2.54		1.76	1.19 to 2.61	
WBC, $\times 10^9$/L*	NA		.0259	NA		.0221
Creatinine clearance, 10-mL/min increase†	0.96	0.93 to 0.99	.0131	0.95	0.92 to 0.99	.0058
Albumin, g/dL*	NA		.0066	NA		.0073
AST, 20-μg/mL increase	1.18	1.09 to 1.27	< .001	1.19	1.09 to 1.28	< .001
No. of study drugs			< .001			< .001
1	Ref			Ref		
≥ 2	1.77	1.47 to 2.13		1.67	1.38 to 2.01	
Agent type			.0086			.0050
Nonbiologic	Ref			Ref		
Biologic	0.55	0.35 to 0.86		0.52	0.33 to 0.82	
Dose						< .001
≤ Highest −3	NA			Ref		
Highest −2	NA			1.60	1.23 to 2.07	
Highest −1	NA			1.53	1.21 to 1.92	
Highest	NA			2.28	1.80 to 2.90	

Abbreviations: ECOG, Eastern Cooperative Oncology Group; OR, odds ratio; NA, not applicable; Ref, referent.
*Restricted cubic splines; OR not applicable.
†Estimated by Cockcroft-Gault equation, capped at 125 mL/minute.

metastatic sites, presence of liver metastasis, or elevated lactate dehydrogenase. None of these factors were significant predictors of drug toxicity in our cohort. Conversely, we found that measures of organ function including bone marrow (low WBC count), kidneys (diminished creatinine clearance), and liver (elevated AST) were important predictors of toxicity despite their lack of prognostic importance. This reinforces a previous finding that SDRT risk is not predicted by prognostic models such as the Royal Marsden Hospital score.[8]

Several commonly used eligibility criteria do not seem to be predictive of SDRT risk, nor are they predictive of survival according to most existing prognostic scores. We found that hemoglobin level, evaluated as either a continuous (using splines) or categorical variable (8 to 9, 9 to 10, and > 10 g/dL) variable, did not predict for SDRT (P = .425 and .268, respectively). Similarly, platelet count was not predictive for SDRT (P = .170). These data suggest that within the range of values represented by patients in this cohort (hemoglobin ≥ 8 g/dL,

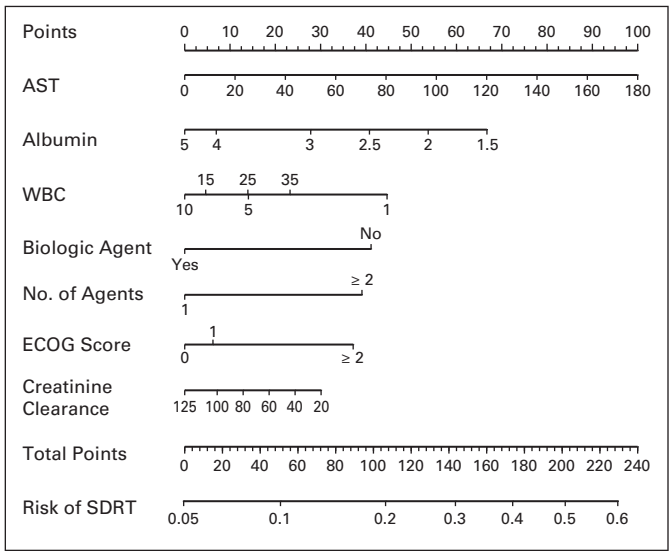

Fig 1. Nomogram for predicting cycle-one serious drug-related toxicity (SDRT) in phase I trials. To calculate probability of SDRT, first determine value for each factor by drawing vertical line from that factor to points scale. Then sum all individual values and draw vertical line from total points scale to risk of SDRT. An electronic tool for calculating risk of SDRT using this nomogram is available online.[18] ECOG, Eastern Cooperative Oncology Group.

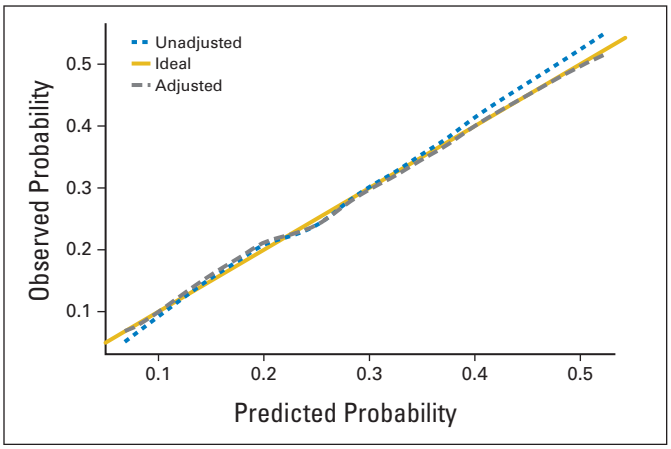

Fig 2. Nomogram model calibration curves. Gold line represents ideal fit, where nomogram-predicted probability (*x*-axis) matches observed probability (*y*-axis). Dashed blue line represents unadjusted calibration accuracy in derivation set and is estimated using LOWESS smoother, relating predicted probabilities to observed binary outcomes. Dashed gray line represents adjusted (bootstrap-corrected) calibration accuracy of derivation set.

platelets $\geq 75 \times 10^9$/L), more stringent parameters do not improve the safety of patients and may unnecessarily limit the population of patients eligible for phase I trials. Similarly, the number of prior systemic therapies was not a significant predictor of SDRT ($P = .720$). Therefore, our data do not support limiting the number of prior systemic therapies in phase I studies unless this is done to look for early signs of drug efficacy.

This study has several important strengths. The model was derived from a large multi-institutional cohort of patients using data from the prospectively maintained CTMS database, which is routinely monitored against source documentation and audited to help ensure data quality. Moreover, the robustness of our final model was demonstrated by internal validation as well as external validation in an independent patient cohort. Model calibration in both cohorts was also good, with the model slightly overestimating the risk in the validation cohort by approximately 5% in certain subsets.

To make the nomogram usable at the time of enrollment, we did not include dose level (which cannot be known in real time relative to highest dose) in the final model, despite the fact that it was significantly associated with SDRT risk. We did investigate whether the omission of dose level biased our final model. As demonstrated by Table 4, the odds ratio and significance level of each covariate remained essentially unchanged whether or not dose level was incorporated into the model. This finding demonstrates that although dose level clearly influences toxicity, it does not meaningfully modify the significance or effect size of the other remaining covariates. As expected, a model incorporating dose level in addition to the other covariates did have an improved overall performance (unadjusted C-index, 0.64). However, we did not feel the modest increase in model performance achieved by incorporating dose level justified substantially curtailing the clinical utility of the nomogram. An alternative nomogram incorporating dose will be made available to researchers who wish to use this tool retrospectively to analyze phase I clinical trial toxicity data. When interpreting these results, it is important to note that clinical as well as statistical reasoning was used to build the nomogram. It is therefore possible that other investigators, given the same data sets, may have made different clinical decisions that result in a slightly different final model because of these choices.

The overall rate of SDRT in the derivation cohort was somewhat higher than has been reported in other contemporary data sets.[8,9] This may, in part, be related to the fact that 63% of patients in the derivation cohort were treated in trials with two study drugs and 57% with a regimen containing a cytotoxic agent. These trial characteristics reflect the tendency of CTEP-sponsored studies to evaluate the combination of novel agents with other investigational or US Food and Drug Administration–approved therapies. The SDRT rate in the validation cohort, which included fewer trials of drug combinations and cytotoxics and more trials involving biologic agents, was lower (13.3% *v* 23.5%) and consistent with other contemporary data sets. Despite these different SDRT rates, the performance of the model in the validation set was excellent, likely because the nomogram accounts for these protocol characteristics, demonstrating the generalizability of the proposed nomogram.

We have shown that there continues to be a significant opportunity to improve the criteria we use to select patients for phase I studies. Using our nomogram, it is possible to identify patients who are at high risk ($\geq$ 30%) for drug toxicity regardless of the dose of study drug they receive. SDRTs that occur in these highly susceptible patients often require the commitment of additional patient and monetary resources to expand dose levels below the MTD. We also found that commonly used eligibility criteria, including strict hemoglobin and platelet parameters, do not seem to improve patient safety, whereas even modest decreases in albumin and creatinine clearance may substantially increase patient risk. In an era of increasing austerity in drug development, it is essential that all efforts be made to prevent premature drug failure that can result from enrolling patients who do not contribute reproducible toxicity data onto our phase I studies. A more sophisticated approach to patient selection, incorporating insights from our nomogram and available prognostic models, should help us improve the safety and scientific validity of our phase I studies without substantially curtailing patient enrollment.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: David M. Hyman, Anne A. Eaton, Mrinal M. Gounder, David R. Spriggs, Percy Ivy, Alexia Iasonos
Administrative support: Gary L. Smith, David R. Spriggs, Percy Ivy
Provision of study materials or patients: Gary L. Smith, Percy Ivy
Collection and assembly of data: David M. Hyman, Anne A. Eaton, Mrinal M. Gounder, Gary L. Smith, Erika G. Pamer, David R. Spriggs, Percy Ivy
Data analysis and interpretation: David M. Hyman, Anne A. Eaton, Mrinal M. Gounder, Martee L. Hensley, David R. Spriggs, Percy Ivy, Alexia Iasonos
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Ho J, Pond GR, Newman C, et al: Barriers in phase I cancer clinical trials referrals and enrollment: Five-year experience at the Princess Margaret Hospital. BMC Cancer 6:263, 2006
2. Karavasilis V, Digue L, Arkenau T, et al: Identification of factors limiting patient recruitment into phase I trials: A study from the Royal Marsden Hospital. Eur J Cancer 44:978-982, 2008
3. Olmos D, A'Hern RP, Marsoni S, et al: Patient selection for oncology phase I trials: A multi-institutional study of prognostic factors. J Clin Oncol 30:996-1004, 2012
4. Arkenau HT, Olmos D, Ang JE, et al: 90-days mortality rate in patients treated within the context of a phase-I trial: How should we identify patients who should not go on trial? Eur J Cancer 44:1536-1540, 2008
5. Arkenau HT, Barriuso J, Olmos D, et al: Prospective validation of a prognostic score to improve patient selection for oncology phase I trials. J Clin Oncol 27:2692-2696, 2009
6. Garrido-Laguna I, Janku F, Vaklavas C, et al: Validation of the royal marsden hospital prognostic score in patients treated in the phase I clinical trials program at the MD Anderson Cancer Center. Cancer 118:1422-1428, 2012
7. Ploquin A, Olmos D, Lacombe D, et al: Prediction of early death among patients enrolled in phase I trials: Development and validation of a new model based on platelet count and albumin. Br J Cancer 107:1025-1030, 2012
8. Molife LR, Alam S, Olmos D, et al: Defining the risk of toxicity in phase I oncology trials of novel molecularly targeted agents: A single centre experience. Ann Oncol 23:1968-1973, 2012
9. Wheler JJ, Tsimberidou AM, Hong DS, et al: Risk of serious toxicity in 1181 patients treated in phase I clinical trials of predominantly targeted anticancer drugs: The M. D. Anderson Cancer Center experience. Ann Oncol 23:1963-1967, 2012
10. Ansher SS, Scharf R: The Cancer Therapy Evaluation Program (CTEP) at the National Cancer Institute: Industry collaborations in new agent

development. Ann N Y Acad Sci 949:333-340, 2001

11. Theradex Systems: NCI-CTMS: Partners since 1982. http://www.theradex.com/CTMS/Default.aspx

12. National Cancer Institute: Common Terminology Criteria for Adverse Events (CTCAE) v.4 2010. http://evs.nci.nih.gov/ftp1/CTCAE/About.html

13. National Cancer Institute: Cancer Therapy Evaluation Program: CTC/CTCAE mapping documents. http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_mapping_docs

14. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-387, 1996

15. International Agency for Research on Cancer: Statistical Methods in Cancer Research. Lyon, France, International Agency for Research on Cancer, 1980

16. Iasonos A, Schrag D, Raj GV, et al: How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 26:1364-1370, 2008

17. Harrell FE: Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY, Springer, 2001

18. Memorial Sloan-Kettering Cancer Center: Prediction tools. http://www.mskcc.org/cancer-care/prediction-tools

19. Ivy SP, Siu LL, Garrett-Mayer E, et al: Approaches to phase 1 clinical trial design focused on safety, efficiency, and selected patient populations: A report from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. Clin Cancer Res 16:1726-1736, 2010

20. Forster MD, Saijo N, Seymour L, et al: Performing phase I clinical trials of anticancer agents: Perspectives from within the European Union and Japan. Clin Cancer Res 16:1737-1744, 2010

21. Ploquin A, Olmos D, Ferté C, et al: Life-expectancy of patients enrolled in phase 1 clinical trials: A systematic review of published prognostic models. Crit Rev Oncol Hematol 83:242-248, 2012

Acknowledgment

We thank Jasmine Pettiford for contributing to data collection for the validation cohort, the U01 investigators for designing and conducting National Cancer Institute Cancer Therapeutics Evaluation Program studies, Theradex for assisting in data analysis, and the patients and families for participating in these studies.

Appendix

Model Building

The primary end point was binary: presence of serious drug-related toxicity (SDRT) in cycle one. Analyses were conducted using logistic regression.[14] Variables included those listed in Table 3. Biologic therapy was defined as study drugs requiring recombinant DNA technology to manufacture. Variables with a P value $\leq .10$ on univariable analysis were considered candidates for the multiple covariate model. To permit nonlinear relationships, continuous variables were modeled with restricted cubic splines with three knots, where the knot locations were determined by the data[17]; in cases where the relationship seemed linear, the nonlinear component of the spline was tested; if it was found to be nonsignificant, the variable was modeled as linear. Categorical variables were grouped based on clinical reasoning. The Breslow-Day test confirmed that the association between candidate parameters and SDRT was similar across agent class (cytotoxic, molecularly targeted, or both; Appendix Table A1).[15] Therefore, one model for all agent types was pursued. The final regression model was chosen based on the clinical and statistical significance of the predictors, following previously published methodology,[16] although statistical significance played an important role. Because of strong correlations between agent class (cytotoxic, molecularly targeted, or both) and biologic study drug (yes or no), these covariates were entered into the model separately. The objective was to create a nomogram that uses baseline pretreatment factors, and thus, treatment (dose received relative to maximum administered) was not included in the final nomogram. However, both nomograms with and without treatment dose are presented for comparison. For each patient, the predicted probability of an SDRT in cycle one was calculated using the final logistic regression model underlying the nomogram. The concordance index (C-index), which is the nonparametric area under the receiver operating curve and is a measure of the ability of the nomogram to discriminate patients with different outcomes, was calculated for these predictions. The variance of the C-indices was estimated using the pROC package in R (http://www.r-project.org; DeLong ER et al: Biometrics 44:837-845, 1988).

Model Validation

The predictive model was validated using 500 bootstrap samples to avoid overfitting. Specifically, a model was built on a bootstrap sample (training set) and then evaluated on the original data set (test set) without modification. Two indices were calculated based on the bootstrap model being evaluated on the bootstrap sample and the original data set. The difference between the two indices was the optimism of the fit. The process was repeated 500 times. The final optimism estimate was calculated as the average of the 500 differences. The difference between the original C-index (unadjusted) based on all the data and the optimism estimate is the unbiased measure of the C-index,[17] which addresses the ability of the nomogram to discriminate among patients if the nomogram were to be used in a new cohort. The calibration of the nomogram, which measures how far predictions are from observed outcomes, was assessed via a calibration plot by plotting the predicted probability of the nomogram for SDRT against the patient-observed or actual probability after using a nonparametric smoothing technique to relate predicted probabilities to observed binary outcomes. Statistical analyses were performed using SAS software (version 9.2; SAS Institute, Cary, NC) and R software (version 2.3.1; http://www.r-project.org), with the ROCR, Design, Hmisc, and pROC libraries.[17]

Table A1. Comparison of ORs by Study Drug Agent Class

Factor	Cytotoxic Agents		Molecularly Targeted Agents		Combination		
	OR	95% CI	OR	95% CI	OR	95% CI	*P**
ECOG performance status (≥ 2 *v* 0-1)	2.64	1.41 to 4.97	1.27	0.72 to 2.24	1.50	0.91 to 2.49	.2094
Hemoglobin (< LLN *v* normal)	1.25	0.85 to 1.84	0.84	0.65 to 1.10	1.21	0.93 to 1.57	.1061
AST (> ULN *v* normal)	1.07	0.70 to 1.62	1.11	0.84 to 1.47	1.49	1.13 to 1.96	.2500
Albumin (< LLN *v* normal)	1.05	0.68 to 1.63	1.14	0.88 to 1.48	1.29	0.29 to 1.68	.7015
WBC (< LLN *v* normal)	1.41	0.72 to 2.75	0.96	0.60 to 1.54	1.90	1.17 to 3.08	.1367
ALT (> ULN *v* normal)	1.07	0.68 to 1.68	1.20	0.89 to 1.62	1.34	0.98 to 1.85	.7034
Bilirubin (> ULN *v* normal)	1.30	0.50 to 3.40	1.30	0.75 to 2.25	1.58	0.96 to 2.58	.8578
Constitutional symptoms (baseline grade ≥ 2 *v* 0-1)	1.29	0.79 to 2.10	1.08	0.79 to 1.47	1.25	0.93 to 1.67	.7415
No. of study drugs (≥ 2 *v* 1)	2.34	1.45 to 3.76	2.06	1.60 to 2.65	NA†		.6472
Creatinine (≥ 60 *v* < 60 mL/min)	0.41	0.25 to 0.68	0.99	0.70 to 1.41	0.71	0.48 to 1.05	.0160
Biologic study drug (yes *v* no)	NA‡		0.47	0.30 to 0.73	0.82	0.30 to 2.21	.3157
Alkaline phosphatase (> ULN *v* normal)	1.12	0.74 to 1.69	1.12	0.87 to 1.44	1.27	0.99 to 1.63	.7581

Abbreviations: ECOG, Eastern Cooperative Oncology Group; LLN, lower limit of normal; NA, not applicable; OR, odds ratio; ULN, upper limit of normal.
*Breslow-Day test.
†All patients had ≥ two study drugs.
‡No patients had biologic study drug.

Table A2. Baseline Patient Clinical Characteristics for External Validation Cohort (n = 234)

Characteristic	No.	%
Primary tumor site		
GI	41	18
Genitourinary	18	8
Thoracic	40	17
Breast	15	6
Gynecologic	45	19
Sarcoma	35	15
Head and neck	22	9
Melanoma and skin	18	8
Sex		
Male	103	44
Female	131	56
Age, years		
Median	60	
Range	21-85	
ECOG performance status		
0	74	32
1	159	68
≥ 2	1	0
Laboratories		
WBC, $\times 10^9$/L		
Median	6.0	
Range	2.3-23.0	
ANC, $\times 10^9$/L		
Median	4.1	
Range	1.3-21.2	
Hemoglobin, g/dL		
Median	12.2	
Range	8.5-16.1	
Platelets, $\times 10^9$/L		
Median	249	
Range	92-657	
Albumin, g/dL		
Median	3.8	
Range	2.2-4.8	
AST, U/L		
Median	26	
Range	12-127	
ALT, U/L		
Median	19.5	
Range	6-167	
Total bilirubin, mg/dL		
Median	0.7	
Range	0.1-1.5	
Alkaline phosphatase, U/L		
Median	89	
Range	31-739	
Creatinine clearance, mL/min*		
Median	92	
Range	24-125	
Study drug agent class		
Molecularly targeted drug	165	71
Cytotoxic	17	7
Cytotoxic and molecularly targeted drug	52	22
Biologic study drug		
Yes	73	31
No	161	69
No. of study drugs		
1	158	68
≥ 2	76	32

Abbreviations: ANC, absolute neutrophil count; ECOG, Eastern Cooperative Oncology Group.
*Estimated by Cockcroft-Gault equation, capped at 125 mL/minute.

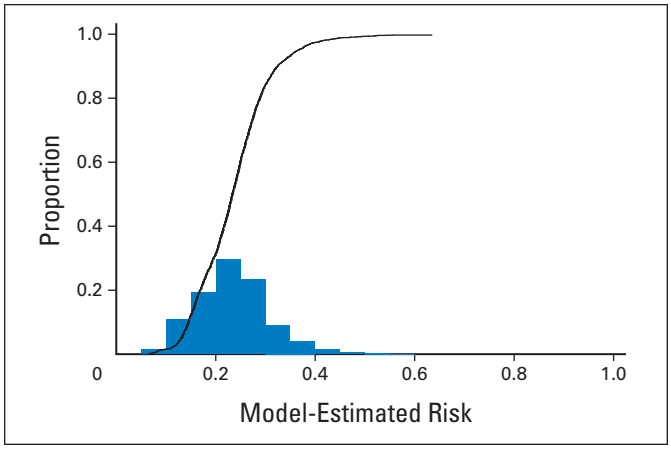

Fig A1. Distribution of model-estimated risk. Histogram of model-estimated risk of cycle-one serious drug-related toxicity in derivation cohort. Line represents proportion of patients with estimated risk at or below given risk (*x*-axis).